

OIIP Workshop

*Web-based Tools and Data Standards for Electronic
Tagging and In situ Datasets*

Fall AGU 2018, Washington DC
Wednesday, 12 December 2018: 08:00 - 12:20

V. Tsontos¹, C. Lam², S. Arms³, N. Quach¹, C. Thompson¹, F. Platt¹, J. Roberts¹, L. McGibbney¹

¹ Jet Propulsion Laboratory, California Institute of Technology

² Large Pelagics Research Center/University of Massachusetts-Boston

³ Unidata/UCAR

OIIP Team



Agenda

- Overview of OIIP project
- In-situ Data Interoperability: Issues & Support Enhancements
- OIIP Data Viewer demo
- ROSETTA Data Converter demo
- Metadata Profile Service overview
- Conclusion & Discussion

Online survey: <https://tinyurl.com/oiip-survey>

Format: Interactive & Feedback corner

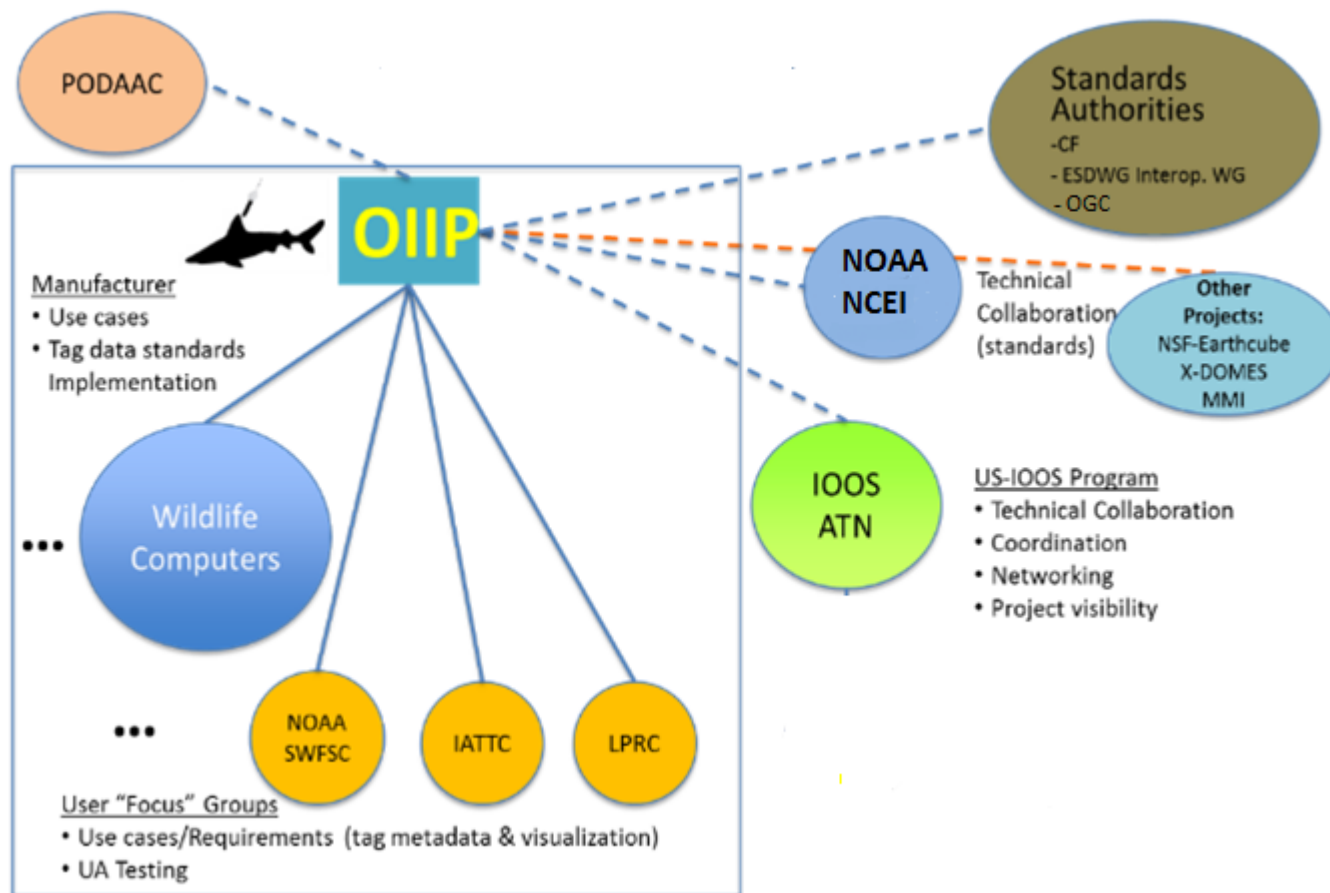
Part 1

OIIP Overview

Motivation & Goals

- **OIIP** is a 2-year NASA/ACCESS15-017 funded technology development project (ended Sept. 2018)
- **Develop improved capacity to support NASA field campaign data** via the enhancement & integration of higher TRL technologies: *NCEI .nc templates*, *ROSETTA*, *THREDDS*, *CMC*, *Tagbase*, *MPS*
- **R&D project addressing some fundamental earth science data informatics challenges:**
 - Need for improved integration of multivariate datasets from diverse observational platforms** in support of NASA science, satellite mission Cal/Val, decision support applications
 - Challenge of *in-situ* data:**
 - inherently diverse, complex, heterogeneous (esp. biological)
 - Acute Interoperability issues:
 - *Ad hoc* data file formats/insufficient adoption of self-describing formats (nc, hdf)
 - Paucity of metadata and/or non-standards compliant metadata
- **Focus on both conventional field campaign & marine animal electronic tagging data** as a representative (but also more challenging) use case
- **Open Source all software components:** source code & Docker containers on Github
Further Info & tutorial/demo videos coming soon on <https://oiip.jpl.nasa.gov>

Community/Stakeholder Engagement



- **Numerous presentations:** Conferences IEEE Oceans, AGU, Oc. Sciences, NASA/ESDSWG, CF2.0 workshop, IOOS-ATN & GEO-MBON F2F meeting (MBARI), 107th OGC-TCP Meeting
- **Upcoming:** IEEE-Oceans-18, Ocean Obs'19 White Paper with NCEI/IOOS

eTag Sensors & Data



SPOT tag

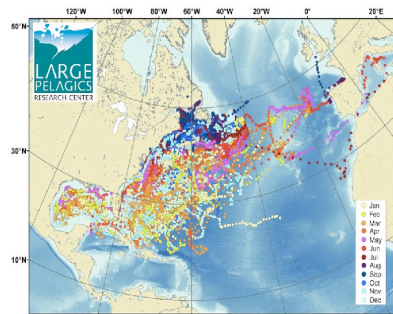


Implantable Archival tag



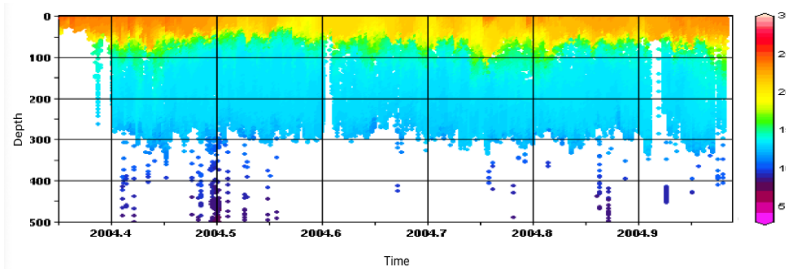
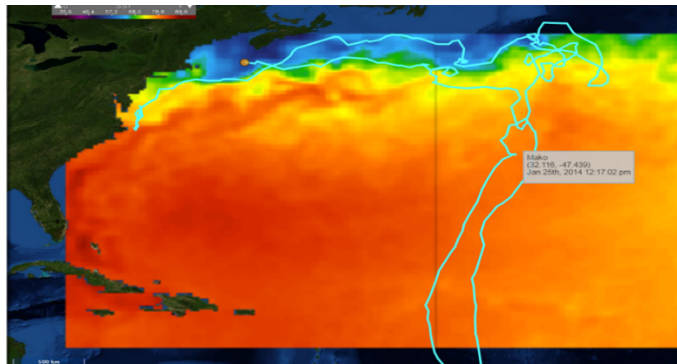
PAT tag on Bluefin Tuna

- Biological “Gliders”
- Horizontally & vertically resolved physical data (*trajectory profiles*)
minimally: light level, pressure/Z, temperature
- Fascinating & Invaluable data: physical oceanography & fisheries management
- Large scale deployments and tagging programs
- IOOS-ATN: *Animal Telemetry Network*

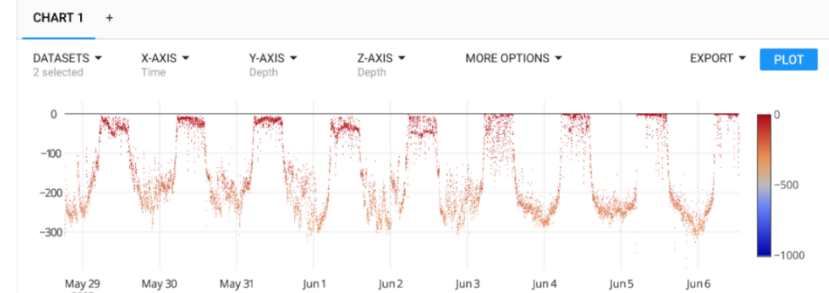
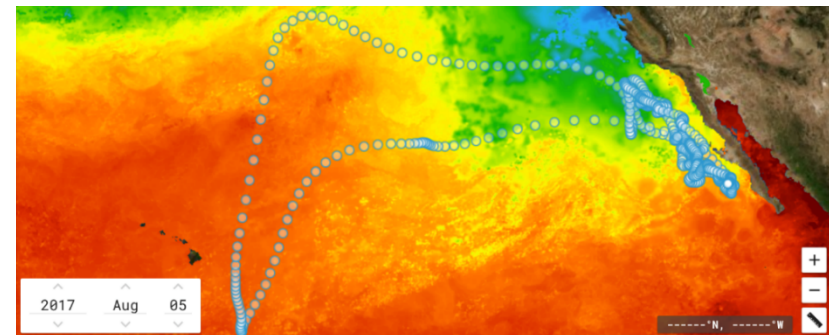


LPRC Bluefin Tuna 1997-2015
(400 tags)

Mako Shark – N. Atlantic, 6 months migration



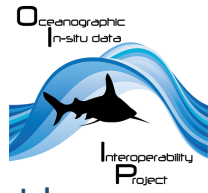
Albacore Tuna– E. Trop. Pacific, 2 years of Archival Data



Part 2

In-situ Data Interoperability

Data Interoperability Problem statements



- There are so many types of data formats for similar instruments from various manufacturers that I use, it is hard to manage, use and synthesize my data [Researcher]
- I wish there was a template specification of what metadata relating to my instrument deployments I should be capturing and packaging with the measurement data. The existing geospatial metadata standards aren't sufficient for actually interpreting the data! [Data Manager]
- I have both an agency and project funding requirement to submit my data for archival at an approved data center but don't have the resources or know how to produce the necessary standard compliant data file for archival [Project Manager]
- A project just submitted 3 years worth of field data for archival as *ad hoc* .CSV files. Its difficult to ingest these into our archive system, make these discoverable, and we cannot easily integrate them into our THREDDS and ERDAP servers either without significant effort [DAAC & IOOS Node]
- Complimentary new, biological sensor platforms fill a current observational gap for tracking dynamic oceanographic features. However, assimilation of these data in our operational model is precluded by the heterogeneity of the data formats and lack of metadata [Modeler]
- I want to easily integrate remote sensing and field observations to produce dynamic visualizations for data exploration and decision support analyses but its way too involved. [Resource Manager]
- We want to be able to support the above community needs but don't sufficiently understand the data interoperability specs or have the time to address this right now. [Instrument Manufacturer]

Data Interoperability: what is it and why is it important?

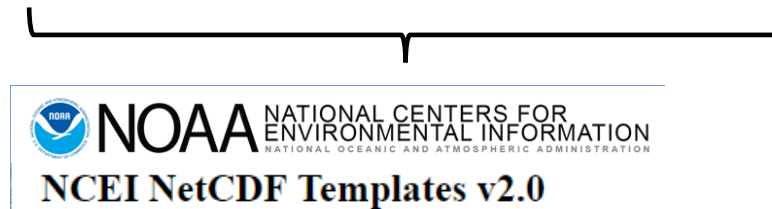
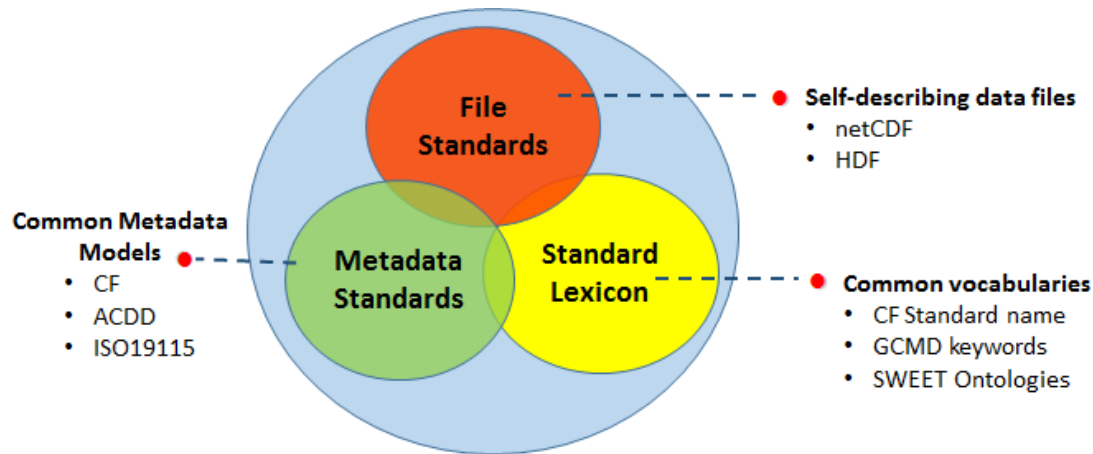
Data Interoperability: The ability of coupled systems to ...

- Communicate and exchange data via common formats & protocols
- Meaningfully interpret and reproducibly act on exchanged data

ie. “*plug-n-play*”, *machine-to-machine*

Syntactic & Semantic Interoperability

Key Ingredients



<https://ioos.noaa.gov/data/contribute-data/common-data-formats/>

“When using netCDF for in situ data, IOOS® data providers should implement feature type [NCEI netCDF Templates](#)”

Significance

- Permit integration with other data types & systems
- Automated data processing & assimilation
- Lower costs of data usage and preservation across the dataset lifecycle

In-situ Data Interoperability Issues

- Highly Heterogeneous native formats (vary b/w Manufacturer/tag types)
- Non-science file formats (invariably CSV)
- Paucity of metadata and/or non-standards compliant metadata
- Lack of support for domain specific metadata
eg. eTag: Instrument, Deployment, Animal, Processing info ...

- Review of NCEI netCDF 2.0 in-situ templates & proposed extensions to support:

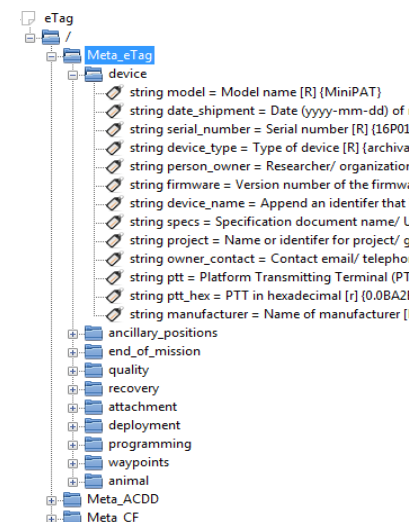
- Rich/community specific metadata, Geolocation uncertainty, Summarized/Non-scalar Data
- Report: [NCEI NETCDF TEMPLATE REVIEW FOR ELECTRONIC TAG DATA SUPPORT](#)
- Presented recommendations to CF2.0 and ESDSWG/DIWG

- Developed a comprehensive, community vetted **inventory & specification of eTag metadata** attributes (130 of which 30 are mandatory) categorized by:
Type (10) & Disposition (3)

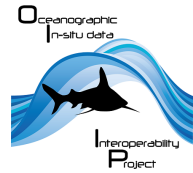
Category/Group	Attribute name	Description	Example	Comments	Necessity	Fish	Shark	Marine	Bird	Notes
device	manufacturer	Name of manufacturer	Wildlife Computers, Microwave Telemetry, Lotek Wireless, Desert Star Systems, CERAS, StarCobb, Sea Mammal Research Unit, Yemco, Loggerhead Instruments, Biotaching Solutions, Little Leonardo, Telemetry etc.		required	1	1	1	1	Some edits to the example by IRTIC
device	model	Model name	MiniPAT		required	1	1	1	1	
device	serial_number	Serial number	16P000		required	1	1	1	1	

- Developed a **framework for packaging such rich metadata attribute sets in .nc4** files utilizing **Groups** to **organize** metadata thematically/hierarchically

Report: [TAG METADATA REVIEW & RECOMMENDATIONS DOCUMENT](#)



Data Standards Work (2/2)



- **eTUFF** (*electronic Tag Universal File Format*)
 - “Normalized” specification in CSV format of any tag data type
 - Flexibly supports enhanced eTag and CF/ACDD/NCEI metadata attributes
 - Shared spec with Wildlife Computers for comment

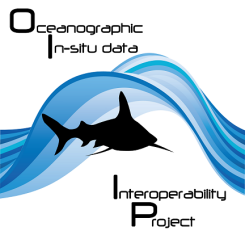
Metadata “header” block

```
// global attributes:
:institution = "IATTC"
:references = "Mar Biol (2010) 157:2625–2642 DOI 10.1007/s00227-010-1524-3"
// etag device attributes:
:person_owner = "Kurt Schaefer"
:owner_contact = "kschaefer@iattc.org"
:device_type = "archival"
:manufacturer = "Wildlife Computers"
:model = "Mk 9"
:serial_number = "590051"
// etag attachment attributes:
:attachment_method = "implant"
// etag deployment attributes:
:datetime_release = "2005-04-15"
:lon_release = "-95.18"
:lat_release = "-1.94"
// etag end of mission attributes:
:end_type = "recaptured"
:end_details = "recovered by fishing fleet"
:date_end = "2009-07-02"
:lon_end = "-83.98"
:lat_end = "-1.45"
```

Tag data block - “normalized” data record representation

```
//Data
//TagID,"DateTime","VariableID","VariableValue","VariableName","VariableSource","VariableUnits"
1,2005-04-15 21:00:00,1,38457.87,"datetime","Proc_WC_PDT",
1,2005-04-16 0:00:00,1,38458.00,"datetime","Proc_WC_PDT",
1,2005-04-16 6:00:00,1,38458.25,"datetime","Proc_WC_PDT",
1,2005-04-16 12:00:00,1,38458.50,"datetime","Proc_WC_PDT",
1,2005-04-16 18:00:00,1,38458.75,"datetime","Proc_WC_PDT",
1,2005-04-17 0:00:00,1,38459.00,"datetime","Proc_WC_PDT",
1,2006-01-16 18:00:00,107,24.00,"PdtDepth03","Proc_WC_PDT","meter"
1,2006-01-17 0:00:00,107,24.00,"PdtDepth03","Proc_WC_PDT","meter"
1,2006-01-17 6:00:00,107,16.00,"PdtDepth03","Proc_WC_PDT","meter"
1,2006-01-17 12:00:00,107,24.00,"PdtDepth03","Proc_WC_PDT","meter"
1,2006-01-17 18:00:00,107,32.00,"PdtDepth03","Proc_WC_PDT","meter"
1,2006-01-18 0:00:00,107,32.00,"PdtDepth03","Proc_WC_PDT","meter"
1,2006-05-07 2:11:00,5,12.00,"depth","Proc_Archival_TempLight","meters"
1,2006-05-07 2:12:00,5,11.00,"depth","Proc_Archival_TempLight","meters"
1,2006-05-07 2:13:00,5,10.00,"depth","Proc_Archival_TempLight","meters"
1,2006-05-07 2:14:00,5,13.50,"depth","Proc_Archival_TempLight","meters"
1,2006-05-07 2:15:00,5,12.00,"depth","Proc_Archival_TempLight","meters"
1,2006-05-07 2:16:00,5,8.00,"depth","Proc_Archival_TempLight","meters"
```

- In lieu of manufacturers natively providing standards compliant netCDF output files ...
Support for eTUFF as a data output option would be an important development:
 - **eliminate** the complexity of the **current, highly variable e-tag data format landscape**
 - ensure **interoperable file production** easily **via tools** such as **Rosetta**



Production of standards compliant data files Critical & Enables good stuff!!!

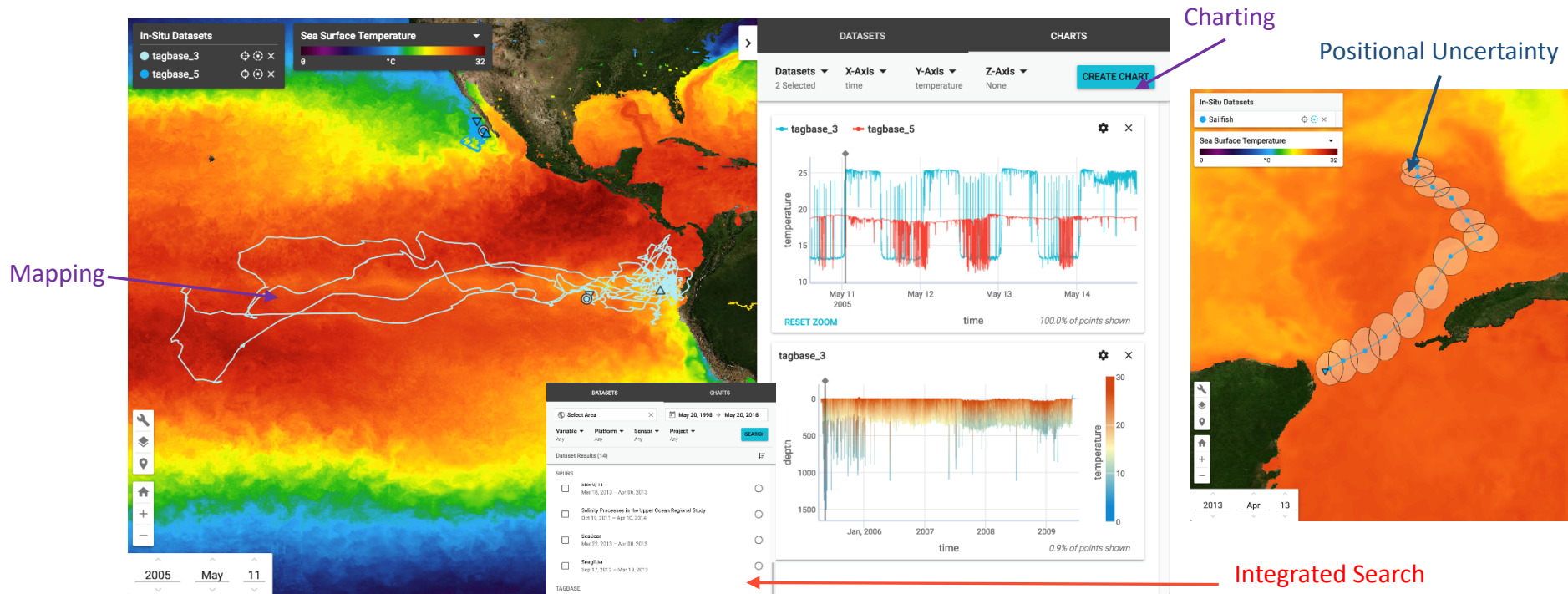
- Ensures data archival & long term preservation at Data Centers
- Data interoperability at the core of operational observing systems
- Facilitates Integration into data access services like THREDDS & ERDAP (enables subsetting and other value-added operations)
- Enables usage in a wide range of applications including:
 - enhanced visualization
 - model assimilation

Part 3

OIIP Data Viewer (CMC)

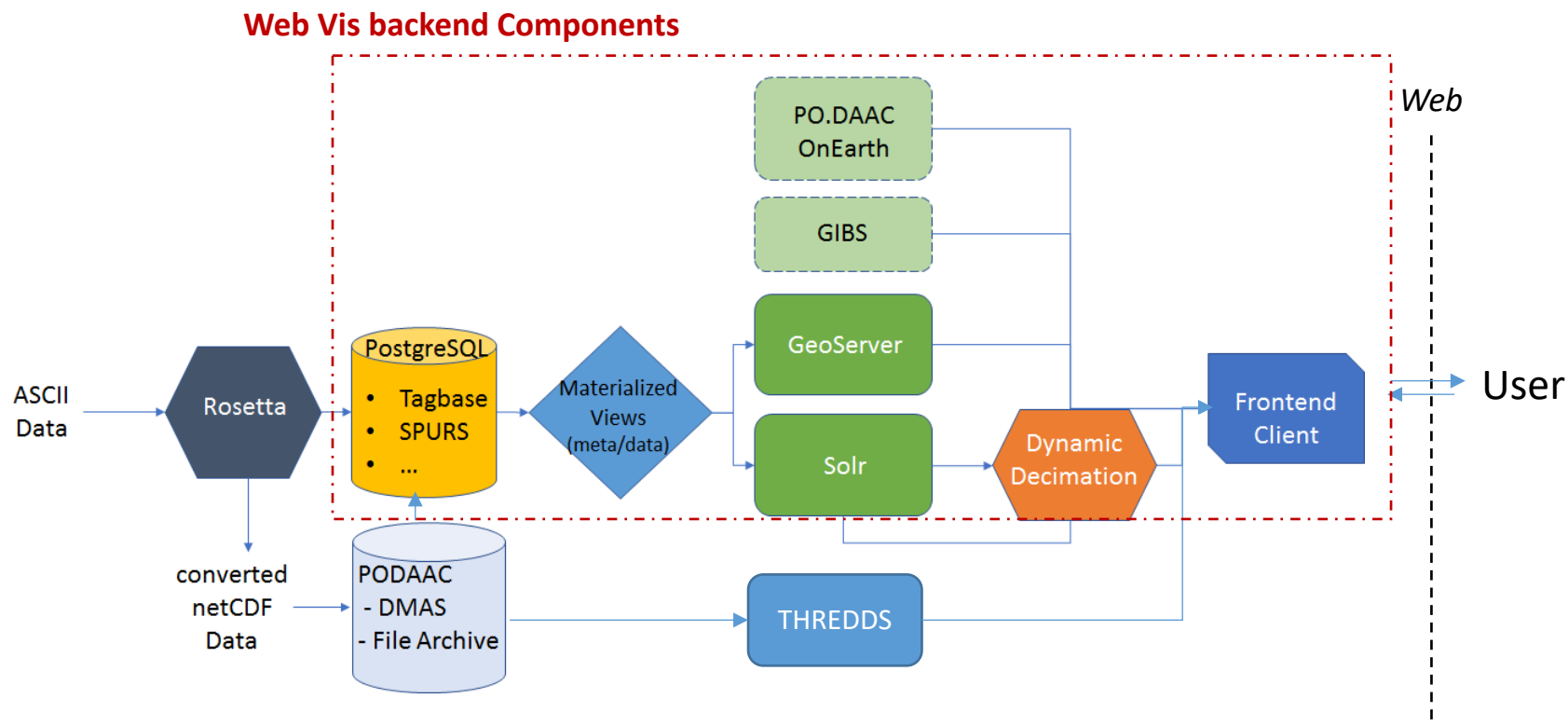
Common Mapping Client (CMC) Web-Visualization

- **Goals:** Develop an intuitive web interface that allows users to:
 - **Integrated visualization** of raster and vector-based *in situ* & satellite datasets (plotting support for all CF-DSG spatial data types)
 - Provide **synchronized horizontal and vertical views of data** and their **evolution over time**
 - **Support high data volume throughput** (data decimation algorithm)
 - **Integrated data search capability** (geospatial with facets for Variable, Project, Platform, Sensor)
- **Approach:** [TAG VISUALIZATION USER CONSULTATION & USE CASES DOCUMENT](#)
- **Open Source Technologies**
 - Front end: JPL-CMC, Open Layers; Back end: Solr, GeoServer, PostgreSQL, GIBS/OnEarth



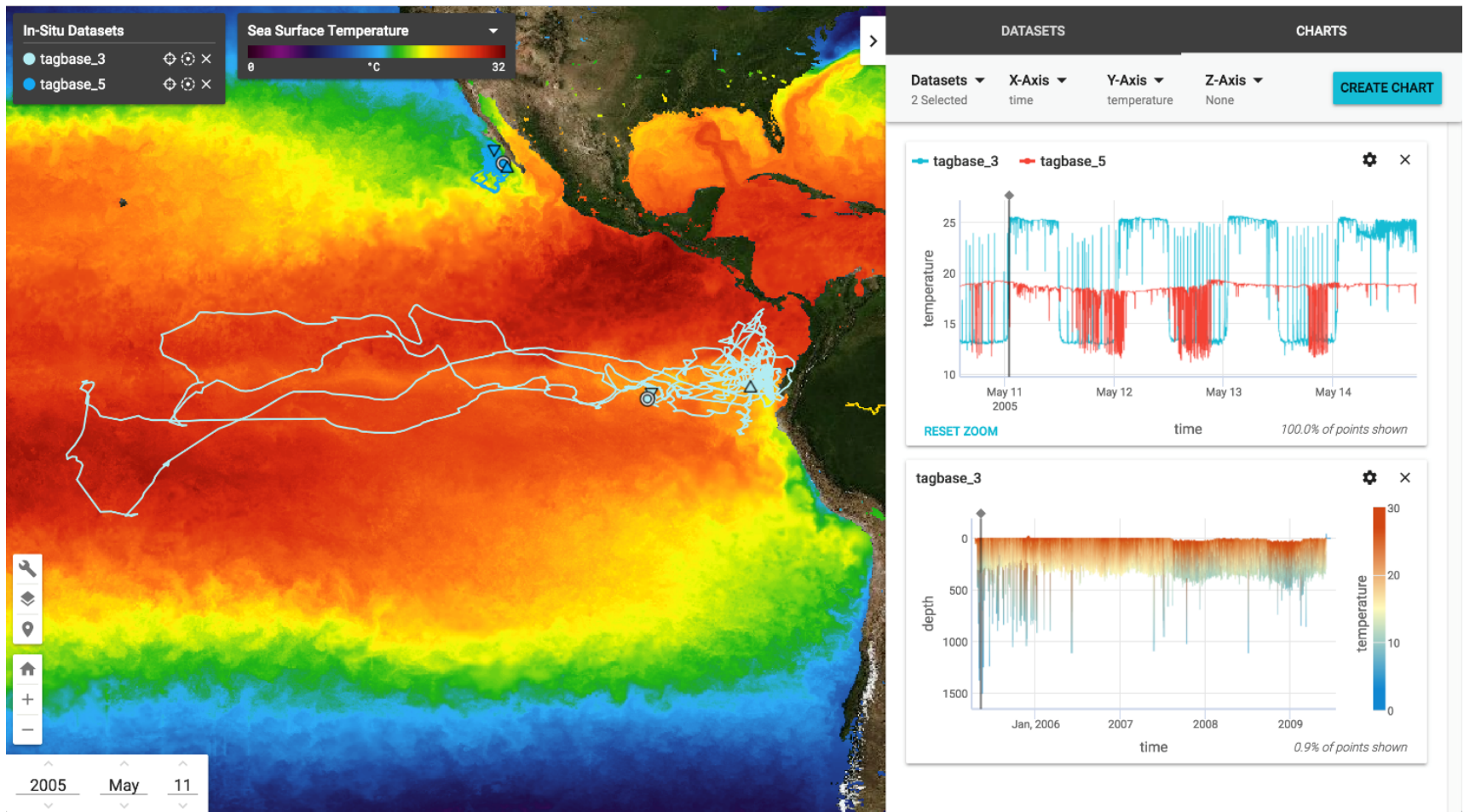
System Integration & Workflow

OIIP components integrate in a test-environment at PODAAC: demonstrate end-to end workflow



OIIP Data Viewer Demo

<https://oiip.jpl.nasa.gov/data-viewer/>



Feedback Corner

<https://tinyurl.com/oiip-survey>

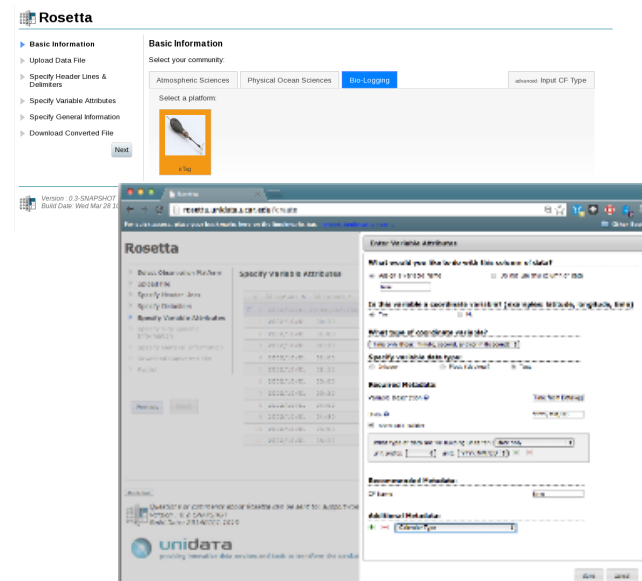
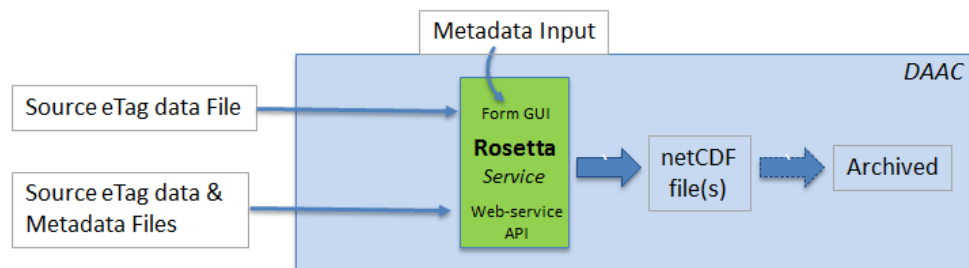
- How did you like the OIIP Data Viewer overall? *(features, design, usefulness)*
- Any aspects you particularly enjoyed or thought were novel?
- Any specific enhancements for the future?

“Standards alone don’t suffice ... Tools facilitate widespread adoption”

Part 4

ROSETTA Data Conversion Tool for Standards Compliance

- Generalized web-based tool for conversion of unstructured/columnar ASCII data files to CF/ACDD compliant netCDF files
- GUI wizard – guided, step-wise process for conversion & augmentation of file metadata by user
- Web service API for bulk/automated conversions



<http://rosetta.unidata.ucar.edu>

OIIP Enhancements:

- ✓ - Extend support for NODC .nc templates & community specific metadata framework
- ✓ - Support for eTUFF and tag metadata profile and workflow
- ✓ - Implement as a RESTful web service interface enabling automated/programmatic data conversions
- ✓ - Integration within OIIP test environment at JPL as a service
- - ROSETTA to be exercised against both eTag & field campaign data
- p - Operational infusion at PODAAC and other data centers

ROSETTA Demo

<http://rosetta.unidata.ucar.edu>



Basic Information

- ▶ Upload Data File
- ▶ Specify Header Lines & Delimiters
- ▶ Specify Variable Attributes
- ▶ Specify General Information
- ▶ Download Converted File

Next

Basic Information

Select your community:

Atmospheric Sciences

Physical Ocean Sciences

Bio-Logging

advanced: Input CF Type

Select a platform:



e Tag



Version : 0.3-SNAPSHOT
Build Date: Wed Mar 28 10:27:30 MDT 2018

Unidata Program Center
support-rosetta@unidata.ucar.edu



Feedback Corner

<https://tinyurl.com/oiip-survey>

- How many of you work with in-situ data?
- Data producers: how challenging is the production of interoperable data files?
- General Impressions on the ROSETTA Concept & Tool?
- Any specific enhancements for the future?
- Would you like to see ROSETTA publicly hosted at a DAAC?

Part 5

Metadata Profile Service (MPS)

Enhanced Support for Granule Metadata

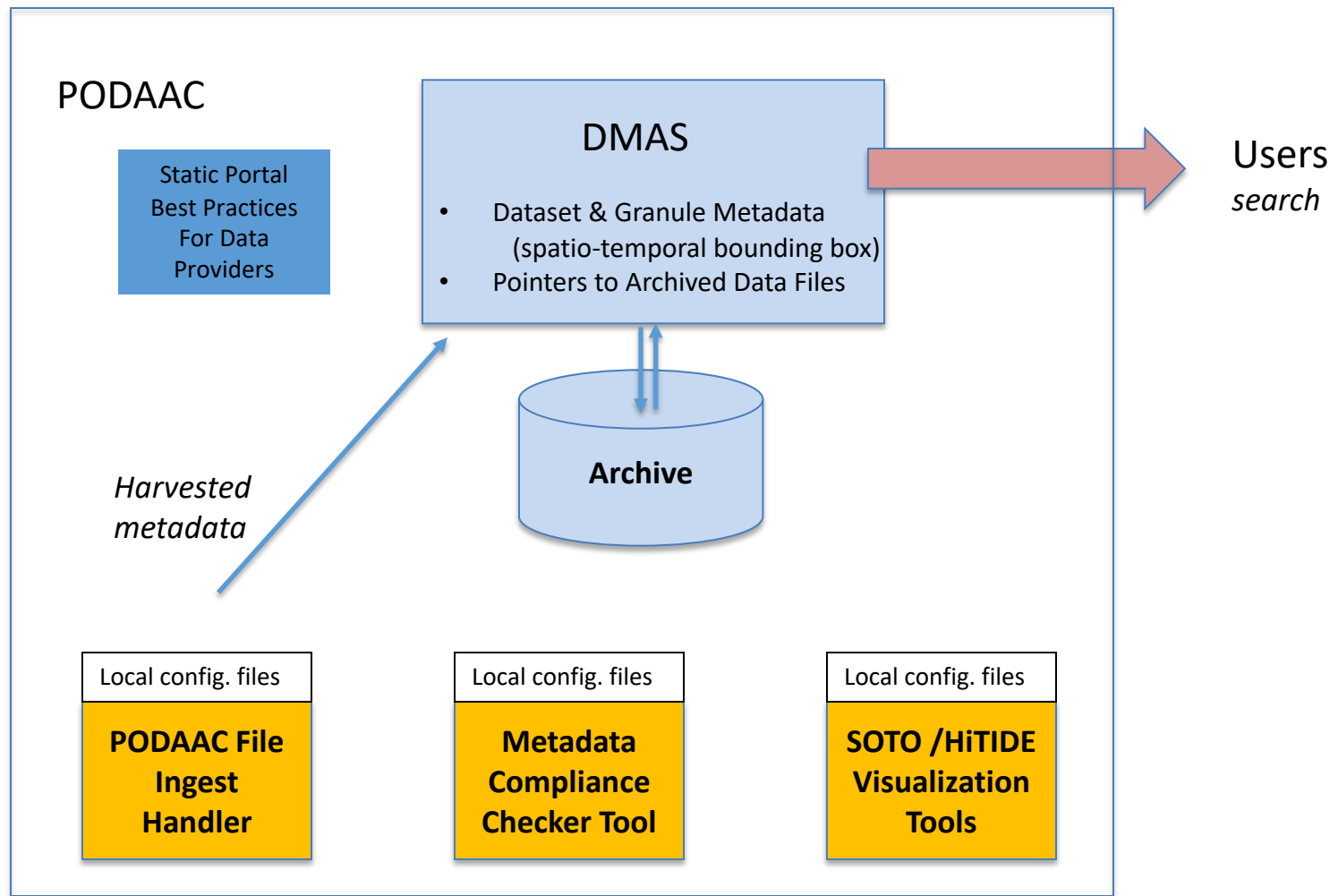
5th Element:

OIIP Deliverable: “Proposed enhancements to DMAS metadata model to better support in-situ”

- In practice involves an analysis of how different “Metadata Profiles” more generally can be modeled and supported within DMAS
 - > profile aware apps
 - > Enhanced search based on Richer metadata

Current Situation at PO.DAAC

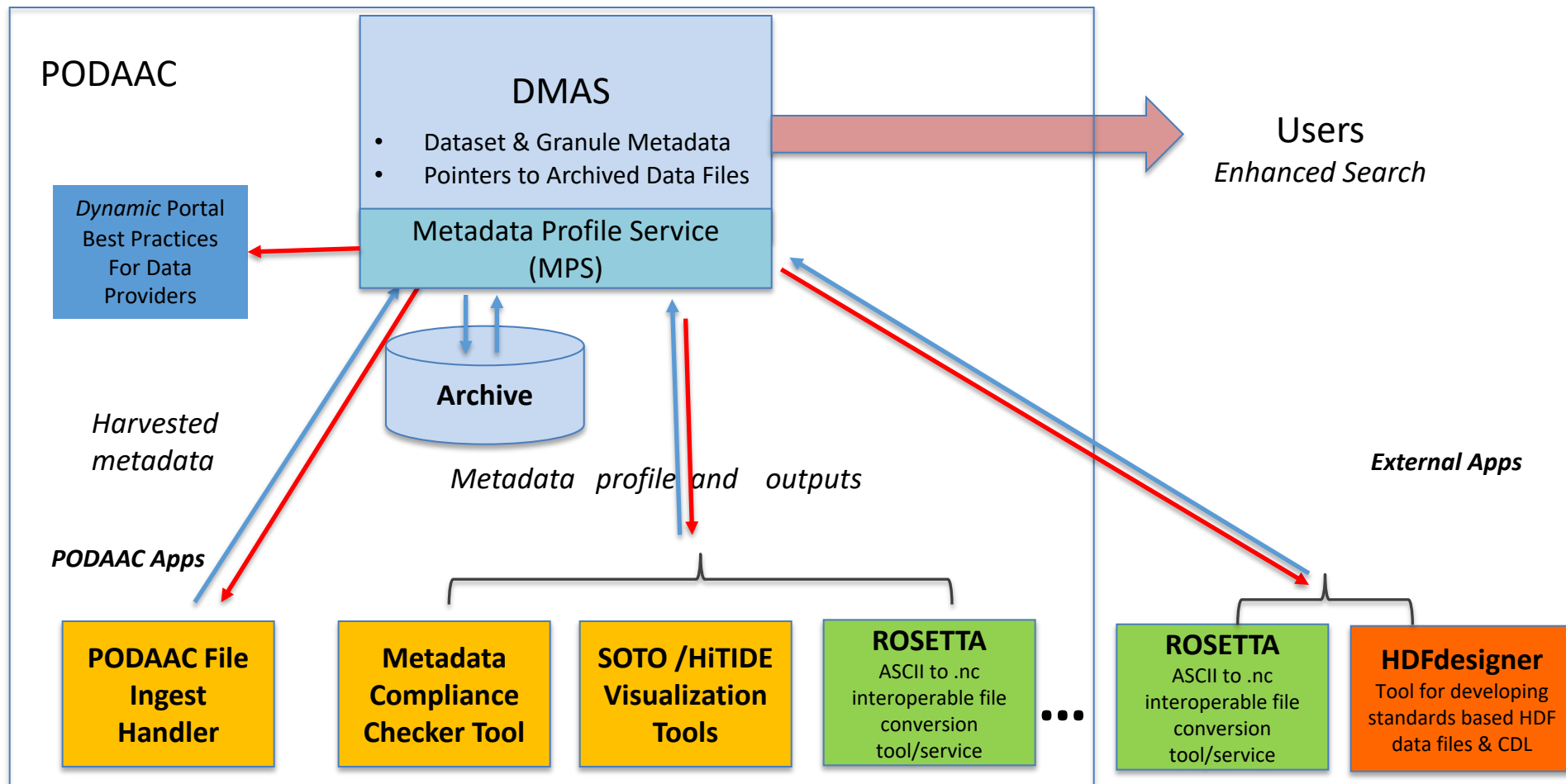
- PODAAC metadata catalog (dataset & granule level)
- Geared towards satellite datasets (relatively constant metadata)
- Granule metadata currently capture limited to bounding box and start/end times
- Support for rich granule level metadata & Richer Search esp. available for in-situ datasets currently lacking
- PODAAC Applications involving metadata harvest/usage are decoupled from DMAS



Metadata Profile Service (MPS)

- Rich metadata support at granule level
- Explicit central registration of metadata profiles (CF, ACDD, GDS, NCEI...)
- ... and traceability to dataset and granule level
- ... for use, via Metadata Profile Service (MPS) coupled with PODAAC and external applications requiring this info

Future ?



MPS Potential

Example Reports/Queries

- What attributes are associated with a given profile and how they are structured within a file (order by scope)?
- What does a given attribute mean, is it required, what format does it take and what are some example values?
- What changes in attributes have occurred between versions for a given profile?
- How does the NCEI v2.0 .nc profile differ from CF1.7 from which it inherits?
- What additional elements does the ARGO profile have relative to NODC? What % of attributes are shared or not
- What combinations of profile standards is eTUFF-nc comprised of? (answer: eTUFF and NODC – which in turn is composed of CF and ACDD)
- What attributes does a given eTag data granule contain and does it conform to the eTUFF-nc profile standards
- Return the attribute values associated with any given granule, dataset and profile
- Return the standard name and units for plotting of a given granule variable by a visualization tool

Feedback Corner

<https://tinyurl.com/oiip-survey>

- Was the MPS concept clear?
- How many think that it is a useful capability to develop?
- Do you see a potential utility & linkage Semantic Web technologies?
- How do you think community vocabularies/ontologies can best be integrated?

Conclusions

- **Data interoperability considerations are vital** because they **enable efficient access & consumption of data** by diverse applications/services **at reduced cost**
- **Acute interoperability problems** associated with in-situ & electronic tagging datasets **is tractable given community convergence around data standards encapsulated by NCEI .nc templates, eTUFF and a conversion tool such as ROSETTA**
- **OIIP has produced a range of useful outputs** for the **Earth Science Data Community**:
 - **standards enhancements & tools addressing in-situ data interoperability issues**
 - **demonstrate the value added of multivariate data integration via enhanced visualization tool**
- **OIIP is continuing to develop** new approaches to maintaining and digesting **community specific metadata profiles**, via the MPS and SIS efforts
- Requesting **community comment over the remaining 3 months** of the project as our tools become available online at **<https://oiip.jpl.nasa.gov>**



Follow us [@OIIP16](https://twitter.com/OIIP16) for news and announcements

Please Take the Online survey: <https://tinyurl.com/oiip-survey>



Comments

Backup Slides